

# BP 网络过拟合现象满足的不确定关系新的改进式<sup>\*</sup>

李祚泳<sup>1)</sup> 彭荔红<sup>2)</sup>

(<sup>1)</sup>成都信息工程学院, 四川, 成都, 610041; <sup>2)</sup>厦门大学环境科学研究中心, 福建, 厦门, 361005)

**摘要** 类比信息传递过程中的一般测不准关系式, 引进表征问题复杂性的函数复相关系数  $R$  和代表网络结构特性的隐节点数  $h$ , 揭示了 BP 网络过拟合现象出现时的网络学习能力与推广能力之间满足的不确定关系式; 通过模拟了 12 种不同类型复杂程度函数的过拟合数值试验, 确定出关系式中的过拟合参数  $p$  的取值范围已缩小为  $1 \times 10^{-5} \sim 5 \times 10^{-4}$ ; 给出应用 BP 网络对给定样本集的训练过程中, 判断出现过拟合现象的方法。

**关键词** BP 网络, 过拟合, 学习能力, 推广能力, 不确定关系。

## NEW IMPROVEMENT OF UNCERTAINTY RELATION SUITED TO OVERFITTING OF BP NEURAL NETWORK<sup>\*</sup>

LI Zuo-Yong<sup>1)</sup> PENG Li-Hong<sup>2)</sup>

(<sup>1)</sup>Chengdu University of Information Technology, Chengdu, Sichuan 610041, China;  
<sup>2)</sup>Research Center of Environment Science, Xiamen University, Xiamen, Fujian 361005, China)

**Abstract** Based on the analogue of general uncertainty relation in information transmitting processes by the introduction of multi-correlative coefficient  $R$  to manifest the complexity of the function and the hidden unit  $h$  to manifest the complexity of the network structure, a general uncertainty relation between the learning ability and the generalization ability suited to overfitting of BP neural network was revealed in the modeling of BP neural network. Tests of numerical simulation for 12 kinds of complicated function were carried out to determine the value distribution ( $1 \times 10^{-5} \sim 5 \times 10^{-4}$ ) of overfitting parameter in the uncertainty relation. Based on the uncertainty relation, the judgement of overfitting in the training process of gived sample sets using BP network was given.

**Key words** BP network, overfitting, learning ability, generalization ability, uncertainty relation.

## 引言

对 BP 网络出现的过拟合现象已有不少研究<sup>[1~6]</sup>。笔者曾在文献[7]中给出了 BP 网络建模出现过拟合现象时的学习能力与推广能力之间应满足的测不准关系式; 并通过数值试验, 确定了测不准关系式中过拟合参数的取值范围为  $1 \times 10^{-6} \sim 1 \times 10^{-3}$ ; 进一步指出参数的取值与问题(逼近函数)的复杂性有关, 而函数的复杂性又可用复相关系数来描述。不过, 由于不同类型的不同复杂程度函数的复相关系数没有严格的区分界限, 因而在实际问题的 BP 网络建模训练过程中, 应用建立的测不准关系式判别过拟合现象是否出现时, 由于判别式中的过拟合参数的取值范围太大, 不易确定, 大大限制了判别式用于指导 BP 网络建模过程中的作用。本文对文

献[7]中的关系式提出了 3 点重要改进: ① 采用网络隐节点个数  $h$  取代原关系式中的网络输入节点个数  $n$ ; ② 在关系式中新引入复相关系数  $R$  表征函数的复杂性; ③ 用相对权值改变量  $\Delta W = \Delta W_1 / \Delta W_2$  取代原关系式中的绝对权值改变量  $\Delta W$ 。通过模拟 12 种不同复杂程度函数的数值试验, 确定出改进后的不确定关系式中的过拟合参数  $p$  的取值范围已大大缩小。在此基础上进一步指出实际问题的 BP 网络建模训练过程中, 为提高网络的推广能力, 防止出现过拟合现象的有效判别方法。

## 1 改进后的 BP 网络过拟合时的学习能力与推广能力之间满足的不确定关系式

在文献[7]中, 类比信息传递过程满足的一般测不准关系式, 提出了 BP 网络建模过程中出现过拟

\* 国家“九·五”重点科技攻关基金(批准号 96-911-08-03)资助项目  
稿件收到日期 2001-06-11, 修改稿收到日期 2001-10-13

\* The project supported by the National “9. 5” Science and Technology Key Foundation of China (No. 96-911-08-03)

Received 2001-06-11, revised 2001-10-13

合现象时,反映学习能力的权值改变量 $\Delta W$ 与表征推广能力的网络对未参与训练样本集的辨识误差 $\Delta_y$ 之间应满足的过拟合测不准关系式

$$|\Delta W| \cdot |\Delta_y| \geq \frac{np}{2\log_2(1+M/N)}, \quad (1)$$

式中 $n$ 为网络输入节点数(样本因子数); $M$ 和 $N$ 分别为网络训练完 $t$ 遍后,训练样本集的网络输出平均归一化值和输出的方均根误差; $p$ 为满足关系式(1)成立的过拟合参数,其取值范围为 $1 \times 10^{-6} \sim 1 \times 10^{-3}$ .

理论分析和应用实践均表明:BP网络的学习能力及推广能力与网络结构(主要指隐层节点数 $h$ )、问题规模(给定样本数 $m$ 和样本因子数 $n$ )和问题复杂性程度(函数的类型)都有关<sup>[5]</sup>.而问题复杂性程度可用函数复相关系数来近似刻划.事实上,问题复杂性不仅与问题的函数类型有关,也与问题规模即样本数 $m$ 和因子数 $n$ 有联系.但关系式(1)的右端的因子数 $n$ 只部分反映了问题的复杂性,而未包含与样本数 $m$ 和函数类型有关的复杂性;关系式(1)右端也没能反映与学习能力和推广能力有关的网络结构参数 $h$ .因此,关系式(1)中的过拟合参数 $p$ 的取值必然随函数类型、样本数 $m$ 和网络结构参数 $h$ 的不同而在较大范围内变化.为了尽可能削弱上述因素对参数 $p$ 的影响,使 $p$ 的取值限制在一个较小范围内变化,对关系式(1)进行以下几点改进:  
① 用代表网络结构的隐节点数 $h$ 取代关系式(1)右端的输入节点数 $n$ ;② 在关系式(1)的右端引进一个能同时表征问题规模和函数类型的复杂程度的复相关系数 $R$ 作因子;③ 用相对权值改变量 $\Delta W = \Delta W_1 / \Delta W_2$ 取代关系式(1)中的绝对权值改变量 $\Delta W$ ,从而关系式(1)改进为

$$|\Delta W| \cdot |\Delta_y| \geq \frac{Rhp}{2\log_2(1+M/N)}, \quad (2)$$

式(2)称为过拟合不确定关系式.上式右端: $R$ 为所逼近函数的复相关系数; $h$ 为构造的BP网络的隐节点数; $M$ 和 $N$ 的意义与式(1)中的意义相同.上式左端: $\Delta W$ 为网络训练过程中出现过拟合现象时,样本集相继训练完两遍后的相对权值改变量,用下列公式计算

$$\Delta W = \Delta W_1 / \Delta W_2, \quad (3)$$

$$\Delta W_1 = \left\{ \sum_{i=1}^n \sum_{j=1}^h \left[ |W_{ij}(t)| \times |V_j(t)| \right] \right\} - \left\{ \sum_{i=1}^n \sum_{j=1}^h \left[ |W_{ij}(t-1)| \times |V_j(t-1)| \right] \right\} / n, \\ \Delta W_2 = \left\{ \sum_{i=1}^n \sum_{j=1}^h \left[ |W_{ij}(t)| \times |V_j(t)| \right] \right\} +$$

$$\left\{ \sum_{i=1}^n \sum_{j=1}^h \left[ |W_{ij}(t-1)| \times |V_j(t-1)| \right] \right\} / 2n,$$

式中 $W_{ij}$ 为网络输入节点 $i$ 与隐节点 $j$ 之间的连接权值; $V_j$ 为隐节点 $j$ 与输出节点 $k$ 之间的连接权值; $t$ 为训练的遍数; $n$ 和 $h$ 分别为网络输入节点数和隐节点数; $\Delta_y$ 为反映推广能力的网络对未参与训练的新样本的辨识误差,采用下式计算

$$|\Delta_y| = \sqrt{\sum_{i=1}^{m'} (O_{ki'} - T_{ki'})^2 / m'}. \quad (4)$$

式中 $m'$ 为未加入训练的新样本的数目, $Q_{ki'}$ 和 $T_{ki'}$ 分别为样本 $i'$ 在输出节点 $k$ 的网络输出值和期望输出值.

## 2 确定过拟合参数 $p$ 值的数值模拟试验

采用常用的3层BP网络数值建模试验.为简化起见,选取网络输出节点个数 $k=1$ 个.数值试验过程中,分别随机选取网络输入节点数和隐节点数的变化范围为 $n=3 \sim 8$ 个和 $h=3 \sim 15$ 个;分别随机选取训练样本数和用作检验推广能力的未参与训练的样本数变化范围为 $m=8 \sim 128$ 个和 $m'=4 \sim 64$ .考虑到应用BP网络对实际问题建模时,通常使用归一化的样本因子值,因此,数值试验中采用各类模拟函数定义域内的因子随机选取值的归一化值作为网络的因子输入值;而训练样本的期望输出值是各类模拟函数计算值的归一化值.本文一共模拟了包括线性函数类、对数函数类、指数函数类、幂函数类、三角函数类、反三角函数类和双曲函数类等12种不同类型复杂程度的函数.对上述各类函数均进行了数百组不同数值试验,得出一致的结果:关系式(2)可视作BP网络过拟合现象出现时,学习能力与推广能力之间应满足的一般不确定性关系式,式中的过拟合参数 $p$ 的取值范围已缩小为 $1 \times 10^{-5} \sim 5 \times 10^{-4}$ ,只有极个别例外.关系式(2)说明:构造一定结构(隐节点数为 $h$ )的3层BP网络对给定样本集进行训练,当训练进行到一定阶段,出现过拟合现象后,若继续训练,虽然可以使 $\Delta W$ 进一步减小,提高了网络的学习能力,但因不等式(2)右端的 $R$ 、 $h$ 、 $p$ 均已固定,而因子 $\log_2 \left[ 1 + \frac{M}{N} \right]$ 改变很小,因此,可以认为不等式右端近似保持不变.为保持不等式(2)成立,左端的 $|\Delta_y|$ 必然增大,即推广能力变差,这就是过拟合现象.

## 3 过拟合不确定关系式用于指导实际建模过程

BP网络应用于实际问题建模的训练过程中,是

否已出现过拟合可借助于过拟合不确定关系式(2)进行判别.大量数值试验结果表明:在网络远未出现过拟合现象的训练阶段,一般说来训练样本的拟合误差  $N$  与未参与训练的样本的辩识误差  $\Delta y$  大致处于同一数量级,即  $N \sim \Delta y$  (当然多数情况的  $|N| < |\Delta y|$ ).此阶段虽然未出现过拟合,但因  $N$  和  $\Delta y$  都较大,因此,若用  $\Delta y$  (或  $N$ ) 代替不等式(2)左端的  $\Delta y$ ,不确定关系式(2)自然成立.随着训练进行到接近或刚开始出现过拟合时,  $N$  与  $\Delta y$  接近相等,若用  $N$  代替式(2)中的  $\Delta y$ ,关系式(2)仍然成立,即有

$$|\Delta W| \cdot |N| \geq \frac{Rhp}{2\log_2(1+M/N)}, \tag{5}$$

由于过拟合参数  $p_0 \in [p_1, p_2] = [10^{-5}, 5 \times 10^{-4}]$ ,故对某具体问题建模时,过拟合参数的限定范围

$$5 \times 10^{-4} \geq \frac{2|\Delta W| \cdot |N| \log_2(1+M/N)}{Rh} \geq 10^{-5} \tag{6}$$

不等式组(6)说明:随着网络训练进行,  $\frac{2|\Delta W| \cdot |N| \log_2(1+M/N)}{Rh}$  计算值逐渐减小,当它小于  $5 \times 10^{-4}$  时,训练就进入了可能出现过拟合的阶段;若继续训练下去,直到计算值大于  $1 \times 10^{-5}$  时,一定已出现过拟合,此时必须停止训练. BP 网络应用于实际问题建模训练过程的过拟合判别操作程序如下:

- a. 由给定的  $m$  个建模训练样本的因子数值和相应的函数值,计算复相关系数  $R^{[8]}$ ;

b. 用 BP 网络对  $m$  个样本训练过程中,不断检验  $p = \frac{2|\Delta W| \cdot |N| \log_2(1+M/N)}{Rh}$  计算值.

若当网络达到指定精度停止训练前,总有  $p > 5 \times 10^{-4}$ ,说明在训练过程中还未处于过拟合阶段,此时不用考虑过拟合;若在达到指定精度停止训练

前,已有  $10^{-4} \leq p < 5 \times 10^{-4}$ ,则认为此时有可能已出现过拟合,就可取停止训练时的  $p_0$  值为过拟合参数值;若达到指定精度停止训练前,就已出现  $10^{-5} \leq p < 10^{-4}$ ,甚至  $p < 10^{-5}$ ,则可认为训练多半早已出现过拟合,尽管模型还未达到指定精度,也应停止训练.表 1 列出了应用 BP 网络于几个实际问题建模训练过程中过拟合判别参数  $p_0$  的取值及有关量计算结果.

4 结论

1. 改进后的 BP 网络过拟合现象出现时,网络的学习能力与推广能力之间满足的过拟合不确定关系式中的过拟合参数  $p$  的取值范围一般为  $1 \times 10^{-5} \sim 5 \times 10^{-4}$ ,已比原有的取值范围缩小了 2 个数量级.

2. 由于复相关系数  $R$  不能完全表征问题的复杂性,因此,既使某些问题有相同的  $R$  值,其复杂性也可能不完全相同,从而使关系式中的过拟合参数  $p$  的取值仍有差异.所以,对于一个实际问题,要准确地确定  $p$  的取值有一定困难,一般只能在  $1 \times 10^{-5} \sim 5 \times 10^{-4}$  范围内选取.

3. 一般说来,若用 BP 网络建立给定样本的分类和识别模型,要求学习能力好是主要的,此时要求  $N$  尽可能小,由判别式(5)知,宜选择接近下限的值作为过拟合参数判定值  $p_0$ ;反之,若是建立 BP 网络的预测模型,要求学习能力和推广能力二者兼顾,则宜选择过拟合参数变化范围的中段值作为过拟合参数判定值  $p_0$ .

4. BP 网络过拟合不确定关系式(2)是普适的,而关系式中的过拟合参数  $p$  的取值范围是大量过拟合数值试验统计结果,能否进一步缩小其取值范围还有待于深入探索.

表 1 几个实际问题的 BP 网络建模的过拟合判别结果

Table 1 The results of the overfitting of BP neural network in modeling of some cases

$n$	$h$	$m$	$R$	$ \Delta W $ ( $\times 10^{-5}$ )	$M$	$N$	$p_0$	$ \Delta W  \cdot  N $ ( $\times 10^{-5}$ )	$Rhp_0/2\log_2(1+M/N)$ ( $\times 10^{-5}$ )
3	6	80	0.8870	43.100	0.6930	0.0735	$1 \times 10^{-4}$	3.1679	7.8663
3	8	42	0.3859	252.763	0.0466	0.1554	$1 \times 10^{-4}$	39.2794	40.7928
4	12	79	0.9280	104.000	0.2820	0.0711	$5 \times 10^{-5}$	7.3944	12.0395
4	8	48	0.9217	200.216	0.3112	0.0797	$1 \times 10^{-4}$	15.9572	16.0687
5	4	79	0.8968	60.234	0.6336	0.1018	$1 \times 10^{-4}$	6.1318	6.2865
5	11	56	0.3460	86.300	0.0327	0.1420	$2 \times 10^{-5}$	12.2546	12.7290
6	12	30	0.8280	25.500	0.1490	0.0440	$1 \times 10^{-5}$	1.1220	2.3288
6	5	63	0.8320	59.900	0.3620	0.1490	$1 \times 10^{-4}$	8.9251	11.6972
7	3	78	0.7770	58.500	0.1180	0.2080	$1 \times 10^{-4}$	12.168	17.9764
7	11	82	0.9089	95.107	0.5443	0.0930	$5 \times 10^{-5}$	8.8449	9.0008

## REFERENCES

- [ 1 ] Levin E. Statistical approach to learning and generalization in layered neural networks. *Proc. IEEE*, 1990, **78**: 1568—1574
- [ 2 ] Wolpert D H. A mathematical theory of generalization. Part I, Part II. *Complex Systems*, 1990, **4**: 151—200, 201—249
- [ 3 ] Moody J O. The dependence identification neural network construction algorithm. *IEEE Trans. on NN*, 1996, **7** (1): 3—15
- [ 4 ] Amari S. Natural gradient works efficiently in learning. *Neural Computation*, 1998, **10**: 251—276
- [ 5 ] YAN Ping-Fan. Some views on the research of multilayer feedforward neural networks. *Acta Electronica Sinica* (阎平凡. 对多层前向神经网络研究的进一步看法. 电子学报, 1992, **27**(5): 82—85
- [ 6 ] PENG H C, GAN Q, WEI Yu. Several realistic approaches to improve the generalization of feedforward neural networks. *Acta Electronica Sinica* (彭汉川, 甘强, 韦钰. 提高前馈神经网络推广能力的若干实际方法. 电子学报, 1998, **26**(4): 116—119
- [ 7 ] LI Zuo-Yong, DENG Xin-Min. Uncertainty relation suited to overfitting of BP neural network. *J. Infrared Millim. Waves* (李祚泳, 邓新民. BP 网络的过拟合现象满足的测不准关系式. 红外与毫米波学报), 2000, **19**(2): 142—144
- [ 8 ] SHI Neng. *Multivariate Analysis Method in Meteorological Statistics Forecasting*. Beijing: Publishing House of Meteorology (施能. 气象统计预报中的多元分析方法. 北京: 气象出版社), 1992, 31—33